

Elemental estimators for the Generalized Extreme Value tail

Allan McRobie
Cambridge University Engineering Department
Trumpington St, Cambridge, CB2 1PZ, UK
fam20@cam.ac.uk

April 17, 2013

Abstract

In a companion paper (McRobie(2013) arXiv:1304.3918), a simple set of ‘elemental’ estimators was presented for the Generalized Pareto tail parameter. Each elemental estimator: involves only three log-spacings; is absolutely unbiased for all values of the tail parameter; is location- and scale-invariant; and is valid for all sample sizes N , even as small as $N = 3$. It was suggested that linear combinations of such elementals could then be used to construct efficient unbiased estimators. In this paper, the analogous mathematical approach is taken to the Generalised Extreme Value (GEV) distribution. The resulting elemental estimators, although not absolutely unbiased, are found to have very small bias, and may thus provide a useful basis for the construction of efficient estimators.

1 Introduction

Together with the Generalized Pareto Distribution (GPD), the Generalized Extreme Value (GEV) distribution is central in extreme value theory. Each distribution has three parameters (μ, σ, ξ) corresponding to location, scale and tail (or shape) respectively, and the estimation of the tail parameter is a problem that has received much attention. McRobie (2013) presented an elegant set of ‘elemental’ estimators for the GPD tail parameter. These estimators: involved three log-spacings; are location- and scale-invariant; and are absolutely unbiased for all tail parameters $-\infty < \xi < \infty$ and all sample sizes $N \geq 3$. The idea was that linear combinations of such elemental estimators could then be constructed which would be efficient, whilst preserving the desirable properties of lack of bias and small sample validity. A variety of linear combinations were considered, and although consistency proofs have yet to be constructed, numerical evidence was presented which suggested that, for distributions within the GPD family, the root mean square error for at least one simple linear combination converged in proportion to $1/\sqrt{N}$ for all ξ .

The idea of this paper is to construct the equivalent elemental estimators for the GEV, using the same mathematical approach as was adopted for the GPD. Here we find that the resulting elementals, rather than being absolutely unbiased, have a very small bias, typically an order of magnitude smaller than the root mean square error. As such, they may thus provide a useful basis within which linear combinations may provide efficient estimators.

In both the GPD and GEV cases, the ultimate aim is of course the application to the wider domain of attraction case, using appropriately chosen maxima from any distribution, rather than from the pure GPD or GEV families. That extension is left for later consideration. Here, only the results for the pure GEV case are presented.

The elemental estimators are illustrated in Figure 1. First, any sample from a GEV is ordered (with X_1 the data maximum and X_N the data minimum). An elemental is then constructed from three log-spacings of the order statistics. The construction takes any pair of non-adjacent order statistics X_I and X_J , and

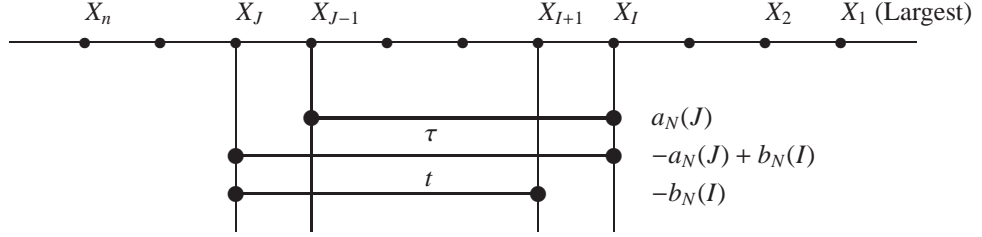


Figure 1: The elemental construction for the GEV case is identical to that for the GPD case, except that different log-spacing weights are used.

involves the log-spacing $\log(X_I - X_J)$ ($J \geq I + 2$) together with two smaller log-spacings $\log(X_{I+1} - X_J)$ and $\log(X_I - X_{J-1})$ that nestle within it. Location- and scale-invariance is ensured by using only the statistics

$$\tau = \frac{X_I - X_{J-1}}{X_I - X_J} \quad \text{and} \quad t = \frac{X_{I+1} - X_J}{X_I - X_J} \quad (1)$$

The resulting elemental estimator takes the form

$$\hat{\xi}_{IJN} = a_N(J) \log \tau - b_N(I) \log t \quad (2)$$

where, as will be shown, the weights $a_N(J)$ and $b_N(I)$ are given by

$$a_N(J) = - \left[\binom{N}{J-1} \sum_{m=0}^{J-1} \binom{J-1}{m} (-1)^m \log(N - (J-1) + m) \right]^{-1} \quad (3)$$

$$b_N(I) = - \left[\binom{N}{I} \sum_{m=0}^I \binom{I}{m} (-1)^m \log(N - I + m) \right]^{-1} \quad (4)$$

These enjoy the property that $b_N(K-1) = a_N(K)$. For the GPD case, the weights did not depend upon the sample size N , but for the GEV case, they do.

The Generalized Extreme Value Distribution arises as the limiting distribution of block maxima (see for example Embrechts et al. (1999)). It has distribution function:

$$F(x) = \begin{cases} \exp\left(-(1 + \xi z)^{-1/\xi}\right) & \text{for } \xi \neq 0 \\ \exp(-\exp(-z)) & \text{for } \xi = 0 \end{cases} \quad \text{where } z = \frac{x - \mu}{\sigma} \quad (5)$$

The parameters μ and ξ can take any value on the real line, whilst σ can be any positive value. For GEVs with positive ξ , the support ($\mu - \sigma/\xi \leq x$) is bounded below but unbounded at the right. For ξ negative, the support is unbounded below but bounded above ($x \leq \mu - \sigma/\xi$). Since the right tail is the tail of interest, positive ξ corresponds to long- or heavy-tailed distributions which are unbounded above, and negative ξ corresponds to “bounded-above” distributions which have a finite right end point.

2 Elemental Estimators for the GEV

The construction of the elemental estimators for the GEV follows that for the GPD case presented in McRobie (2013). The detailed construction is given in Appendix 1 here. In outline, a sample of size N

is drawn from a member of the pure GEV family. The sample is then ordered, with X_1 the sample maximum and X_N the sample minimum. The location- and scale-invariant statistics τ and t are then constructed, these being ratios of data spacings as defined in Fig. 1 and Eqn 1.

The expected values of $\log \tau$ and $\log t$ are then determined for the two cases where ξ is positive or negative. These integrals are expressed via the Probability Integral Transform as integrals of the uniformly-distributed distribution function $F(\mathbf{X})$ over the N -dimensional unit simplex corresponding to the ordered sample.

These integrals readily decompose into a simple part (of the form $\xi \langle \log(-\log F) \rangle$) and a complicated part (of the form $\langle \log(1 - \phi^{|\xi|}) \rangle$). The simple part leads to a term proportional to ξ , and forms the core of the estimator. For the GPD, the elemental combination of $\langle \log \tau \rangle$ and $\langle \log t \rangle$ whose simple terms deliver the estimate ξ has the pleasing property that the complicated terms cancel exactly at all ξ (such that the complicated integrals do not need to be evaluated explicitly). However, in the GEV case, the elemental combination whose simple parts deliver the estimate ξ does not cancel the complicated parts. However, as will be demonstrated numerically, the residual error in this non-cancellation is typically much smaller than the variability.

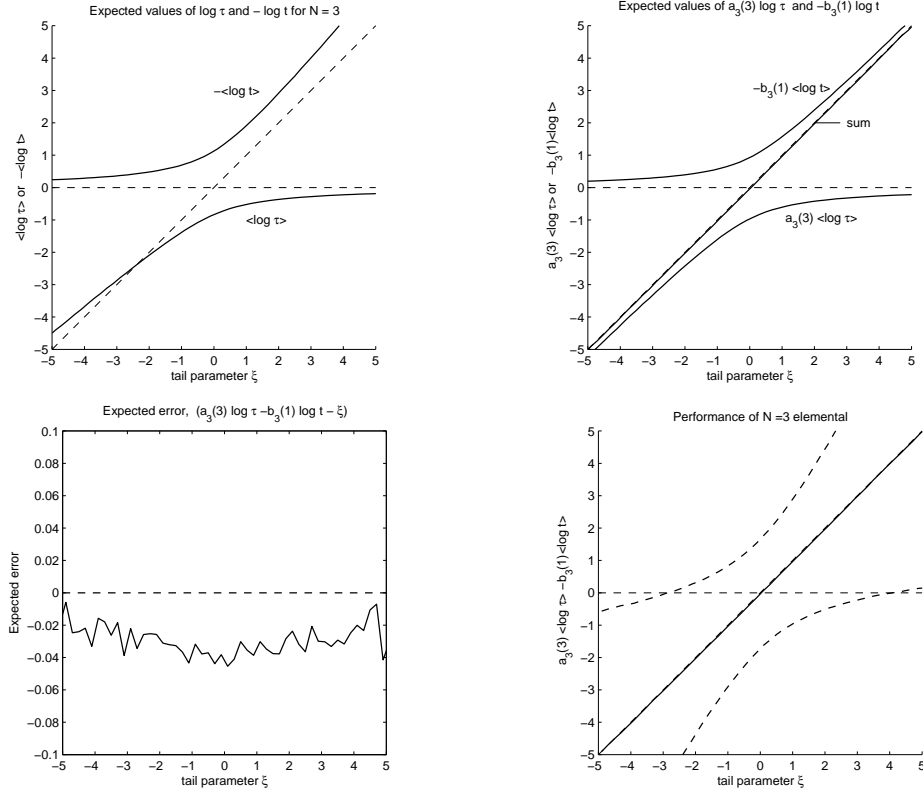


Figure 2: The elemental construction for $N = 3$. The first figure shows how the expected values of $\log \tau$ and $-\log t$ have approximately hyperbolic form, and the second figure shows how scaling these by $a_N(J)$ and $b_N(I)$ makes each asymptotic to the diagonal for large $|\xi|$. Their sum then lies very close to the diagonal, but not exactly. The third figure shows the small bias, and the fourth shows the standard deviation of the estimates.

Although the construction thus far has been developed by simply mimicking the procedure adopted for the GPD elements, Fig. 2 shows how the construction is actually a natural one for the GEV per se. For

$N = 3$, the first figure shows how the expected values of $\log \tau$ and $-\log t$ have approximately hyperbolic form, asymptoting to zero at one end and to some linear slope at the other. The second figure shows how scaling these by $a_N(J) = a_3(3) = -1/(3 \log(3/4))$ and $b_N(I) = b_3(1) = -1/(3 \log(2/3))$ makes each asymptotic to the diagonal for large $|\xi|$. The elemental estimator is the summation, and this lies very close to the diagonal as desired, but not exactly so. It is asymptotic to the diagonal for large $|\xi|$ (as may be expected since the expected values of the complicated terms $\log(1 - \phi^{|\xi|})$ tend to $\log 1 = 0$ for $|\xi|$ large), and the small but non-zero bias for intermediate values of $|\xi|$ is shown in the third figure. The standard deviations of the estimator are shown in the fourth figure, from whence it can be seen that, at worst (near $\xi = 0$), the bias is some 50 times smaller than the standard deviation. These figures were produced numerically, using 250,000 samples of size $N = 3$ at each value of ξ . Although the figures correspond to the single elemental available at $N = 3$, similar figures apply for any elemental at any N .

In summary, the elementals for the GEV are constructed by choosing weights $a_N(J)$ and $b_N(I)$ which guarantee lack of bias at large $|\xi|$. As stated, one hopes that the bias at any intermediate value of ξ is small, and happily it is typically much smaller than the variability there (and even if larger intermediate deviations were to occur for larger N , it is clear that linear combinations of elementals could be constructed such as to manage the overall deviation).

3 Evaluation of the Coefficients

Although the elemental coefficients $a_N(J)$ and $b_N(I)$ are given by a comparatively simple summation (Eqns. 4 and 4) involving binomial coefficients and logarithms, the evaluation becomes fraught with numerical errors for $N \gtrsim 25$ in 32 bit precision computation. The numerical oscillations are illustrated in Fig. 3. That this is a numerical instability rather than true behaviour is noted from the fact that if one makes less effort to calculate the factorials carefully then the instability occurs at lower N . Even when care is taken, and the binomial coefficients are calculated via the standard procedure of using logarithms of gamma functions, there is only marginal improvement in numerical performance. The problem arises because the summations in Eqns. 3 and 4 sum close to zero, but the individual elements in the sums - the logarithms weighted by the binomial coefficients - may be large. Repeated addition and subtraction of these, via the $(-1)^m$ factor, means that small rounding errors accumulate and eventually dominate.

Before describing how these numerical instabilities can be resolved, we note that the coefficients $a_N(J)$ follow trivially from the $b_N(I)$ coefficients using the relation $a_N(J) = b_N(J - 1)$ (with the nuance that $b_N(I)$ must also be evaluated at the meaningless value $I = N - 1$ in order to capture the coefficient $a_N(N)$ which would not otherwise be computed). We thus proceed to describe only the calculation of the $b_N(I)$.

We define $\beta_N(I) = 1/b_N(I)$, such that β is the weighted summation and its reciprocal b is the desired coefficient. It follows from Eqn. 4 that the coefficients β can be computed using the recursion relation

$$\beta_N(I + 1) = \frac{N}{I + 1} \beta_{N-1}(I) - \frac{N - I}{I + 1} \beta_N(I) \quad (6)$$

This is seeded by the $I = 1$ top row

$$\beta_N(1) = N \log \left(1 - \frac{1}{N} \right) \quad (7)$$

Each β coefficient in row $I + 1$ is thus computed from a weighted sum of two terms above it in row I , much akin to a standard Pascal's Triangle construction. This elegant construction avoids all need for explicit computation of factorials or logarithms of gamma functions. However, even this algorithm does not remove the numerical inaccuracies that arise for $N \gtrsim 25$, thus an additional approach is required for large N .

It can be readily seen that, to first order (ignoring terms of order $1/N$), the recursion formula reduces to

$$\beta_{N-1}(I) = (1 - x) \beta_N(I) + x \beta_N(I + 1) \quad (8)$$

where $x = I/N$. This is simple linear interpolation amongst the three coefficients. To first order, then, the approximate recursion formula Eqn. 8 states that the three values of β lie on a straight line when plotted as a function of x . This suggests that there is some underlying curve to which the β values are converging.

Numerical evidence presented in Fig. 3 b) suggests that

$$\lim_{N \rightarrow \infty} \left[N \binom{N}{I} \sum_{m=0}^I \binom{I}{m} (-1)^m \log(N - I + m) \right] = \frac{1}{(1-x) \log(1-x)} \quad \text{where } x = \frac{I}{N} \quad (9)$$

such that $b_N(I)/N$ converges to $f(x) = -(1-x) \log(1-x)$ as N increases. An analytical demonstration of this is presented in Appendix 2.

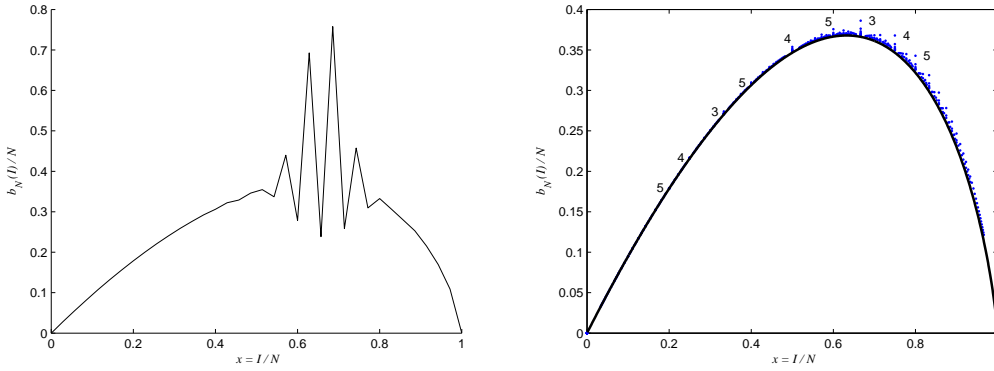


Figure 3: Calculation of the elemental coefficients $b_N(I)$. The left hand diagram illustrates erroneous numerical oscillations from the recursion formula, clearly visible for $N = 35$. The right hand diagram shows the convergence of $b_N(I)/N$ (shown as dots for $N = 3$ to 30) to $f(x) = -(1-x) \log(1-x)$ (shown solid), where $x = I/N$.

Further numerical analysis suggests that, for finite N , there is a better approximation

$$\frac{b_N(I)}{N} \approx -(1-x) \log(1-x) - \frac{x}{12N} \log(1-x) \quad (10)$$

The numerical evidence for this is shown in Fig. 4 where, for N up to 30, the actual values of $b_N(I)/N$ are plotted as points and the approximations are plotted using small circular markers. In all cases, even for $N = 3$, each actual value lies well within the corresponding circular marker. The relative error $(b_N^{approx} - b_N)/b_N$ is shown in Fig. 5, showing convergence at any I and a better than 1 % accuracy for all $N \leq 25$ (and in most cases shown, the accuracy is considerably better).

In summary, for $N \leq 25$, the $b_N(I)$ are computed by the recursion formula Eqn. 6, and for $N > 25$ we use the approximation for $b_N(I)$ contained within Eqn. 10. The corresponding values of $a_N(J)$ are computed from the relation that $a_N(K) = b_N(K-1)$.

For the GPD case McRobie (2013), the elemental coefficients may be succinctly expressed as

$$a_N^{GPD}(J) = J - 1 \quad \text{and} \quad b_N^{GPD}(I) = I \quad (11)$$

Unlike the GEV case, these are independent of N . Fig. 5b) shows that the GEV coefficients converge towards the GPD weightings for the extreme tails (i.e. for x small), much as may be expected.

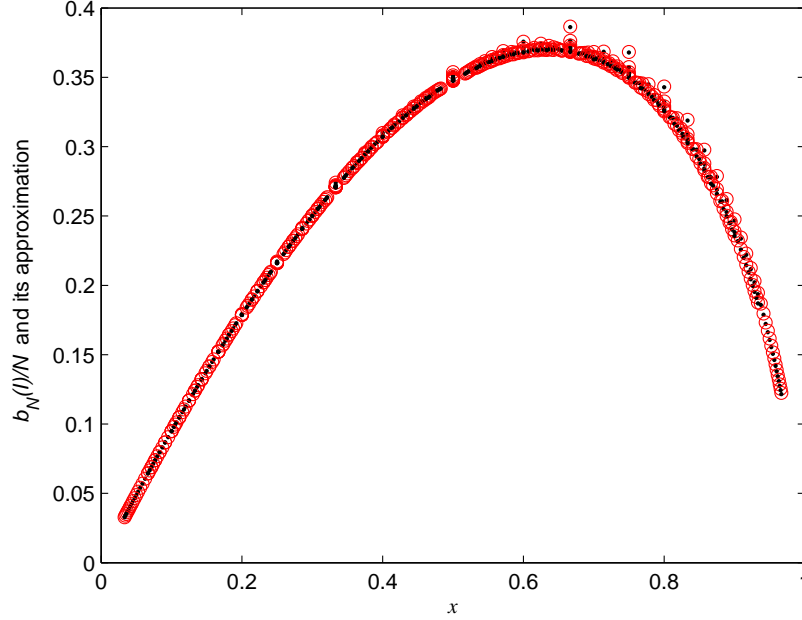


Figure 4: The approximation of $b_N(I)/N$ by Equation 10. The $b_N(I)/N$ values are plotted as points and the approximations are plotted using small circular markers.

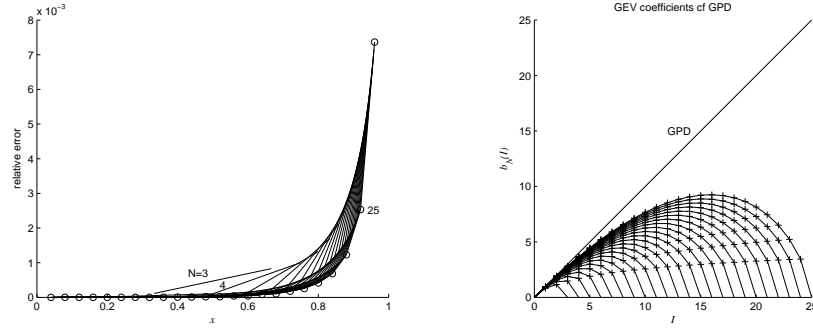


Figure 5: Left: the relative error involved in the approximation of $b_N(I)/N$ by Equation 10. Right: a comparison of the elemental coefficients $b_N(I)$ of the GEV with those of the GPD.

4 Performance of the elementals

Fig. 6a) shows the performance (mean and mean \pm std. dev) of each of the fifteen elemental estimators available at $N = 7$ when applied to samples drawn from a pure GEV. It can be seen that each elemental is very close to being unbiased. It can also be seen that some standard deviations are large. The biases are plotted in Fig. 6b). Also plotted in each of these figures is the performance of the unit sum linear combination which gives equal weight to each of the fifteen elementals. As may be expected, the linear combination - like the elementals - is almost unbiased, and generally has a considerably smaller standard deviation than the individual elementals.

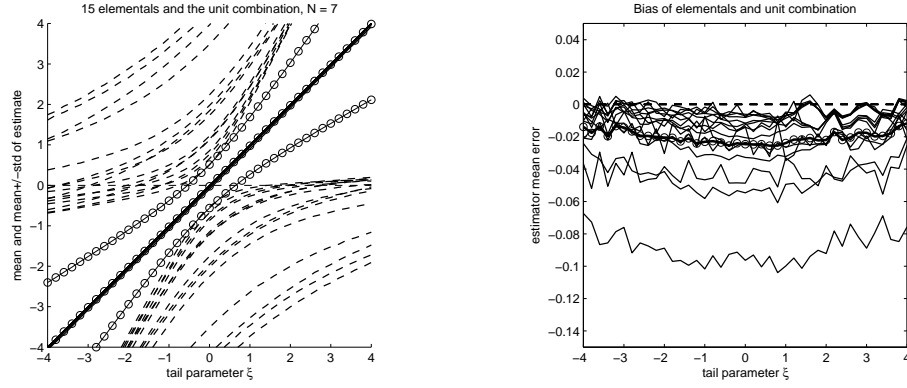


Figure 6: The left hand diagram shows the performance (mean, and mean \pm standard deviation) of each of the fifteen elementals available for samples of size $N = 7$. Also shown (circles) is the performance of the unit sum linear combination which gives equal weight to each elemental. The right hand diagram shows the small bias of the elementals and the linear combination. Plots based on 500,000 samples at each ξ .

The question of which linear combination is in some sense optimal is addressed in Fig. 7 where seven different linear combinations of the individual elementals are considered. The weights in each linear combination have unit sum. The first combination gives equal weight to each elemental. The next three combinations resemble linear triangular basis function which - prior to normalisation - have unit weight at one corner of the upper triangular matrix of possible I and J values, and zero weight at the other two corners. The weights are thus proportional to $N - J + 1$, $J - 1 - I$ and I respectively. The final three combinations are sums of two of the previous three, such that - pre-normalisation - they have unit weight at two corners and zero weight at the remaining corner. There is nothing special about this choice of combinations, other than that the choice serves as a useful starting point.

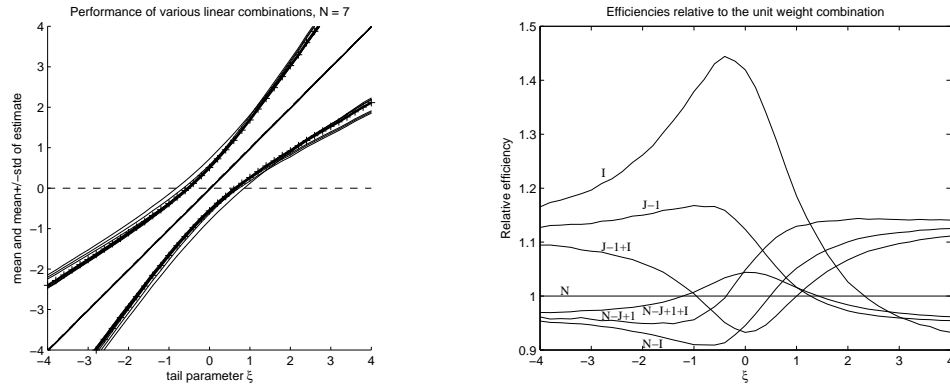


Figure 7: The left hand diagram shows the performance (mean (solid), and mean \pm standard deviation (dashed)) of seven different linear combinations of the fifteen elementals available for samples of size $N = 7$. The equal weight combination is marked with + signs. The right hand diagram shows the relative efficiency (ratio of root mean square errors) of each combination to the equal weight combination (denoted N). Plots based on 100,000 samples at each ξ .

Fig. 7a) shows that the combinations perform similarly. By plotting the ratio of root mean square errors,

Fig. 7b) makes clearer the small differences in performance showing the efficiency of each combination relative to the equal weight combination. It can be seen that some combinations perform better in some parameter regimes, and that no combination is optimal for all ξ . Whilst one could consider constructing some form of adaptive estimator - a first estimate of ξ being used to select the linear combination that is efficient in that region - at this stage of the analysis, we simply proceed with any convenient choice of combination, such as that with equal weights, or - to mirror the choice used in the GPD case - the combination whose weights are proportional to $N - J + 1$.

5 Consistency

No attempt is made here at providing an analytical proof of consistency for any combination. One reason for this is that small and moderately-sized data sets are often of practical interest, and the usefulness of a consistency proof requiring very large sample sizes is then debatable.

Instead, numerical evidence is provided to demonstrate that root mean square errors of some standard linear combinations appear to converge sensibly as sample sizes increase (with samples drawn from pure GEVs). Fig. 8a) shows how root mean square errors decrease at a selection of parameter values ξ as sample size grows from $N = 3$ to 30, for the combination whose weights are proportional to $N - J + 1$. This evidence is extended in Fig. 8b) to samples of size $N = 1000$. The apparent linearity of the graphs on the chosen axes suggests that root mean square error decays as $1/\sqrt{N}$.

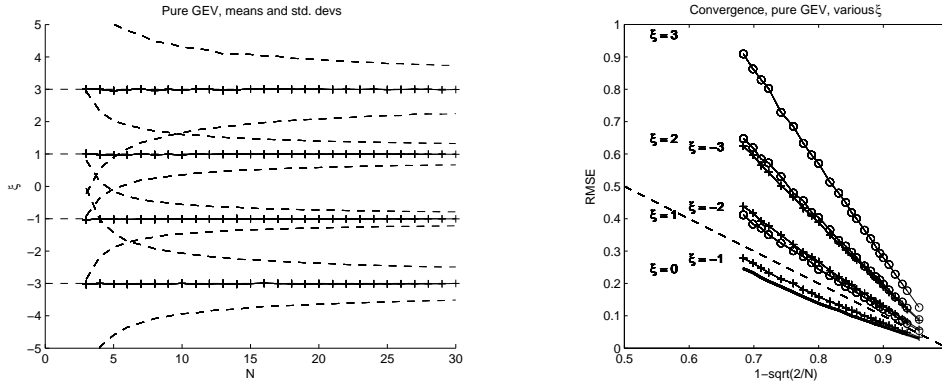


Figure 8: The left hand diagram shows the performance (mean, and mean \pm standard deviation) of the elemental combination with weights proportional to $N - J + 1$, for sample sizes from $N = 3$ to 30. Means are shown (+) with deviations dashed. The right hand diagram extends this to samples sizes up to $N = 1000$. Cases with ξ negative are shown (+) and those with ξ positive are shown (o). The y-axis is the root mean square error (based on 10,000 samples) and the x-axis plots $1 - \sqrt{2/N}$. The tendency towards linearity under increasing N at each ξ suggests that there is consistency, with convergence at a rate proportional to $1/\sqrt{N}$.

6 Comparison with Maximum Likelihood

Maximum Likelihood Estimation (MLE) is a standard tool for parameter estimation, and Embrechts et al. (1999) state that the numerical calculation of ML estimates for the GEV “poses no serious problem *in principle*”. Even with their italicised qualification, this is perhaps optimistic. For example, it is well-known that MLE applied to the GEV requires regularity conditions that do not hold when $\xi \leq -1/2$ (Smith

(1987)). Moreover, for any data set, there are always parameter sets wherein the likelihood is arbitrarily large (i.e. the likelihood is infinite everywhere on the boundary surface $1 + \xi(x_{\max} - \mu)/\sigma = 0$ for $\xi < -1$, and is arbitrarily large in open neighbourhoods adjacent to this). Castillo and Daoudi (2009) also show that there are occasions where the likelihood function for the two-parameter GPD has no local maximum.

Fig. 9 shows the estimates obtained via Maximum Likelihood and via the equally-weighted combination of elementals for 10,000 samples of size $N = 7$ drawn from a GEV. The parameters are $(\mu, \sigma) = (0, 1)$ and ξ is uniform random over $[-10, 10]$. The Maximum Likelihood algorithm used is the *gevfit* function within the Statistics Toolbox of Matlab (MathWorks (2012)). The iterative ML analysis took considerably longer to run than the straight-forward elemental evaluation, and the unusual pattern of the ML results suggests that the algorithm experienced some numerical difficulties. The elemental results, on the other hand, are as one may expect - showing unstructured scatter about a diagonal mean.

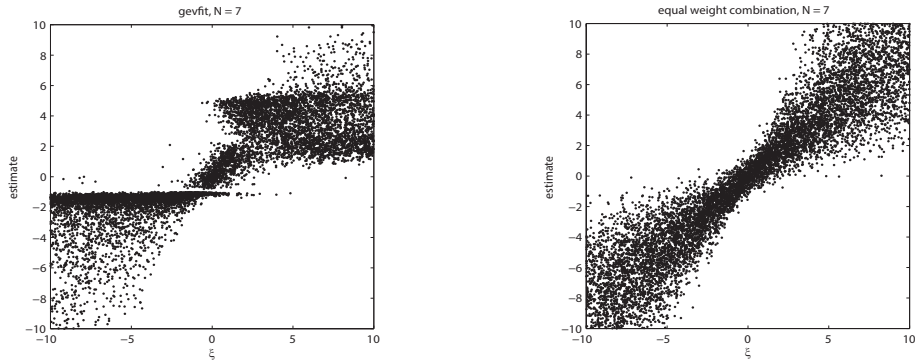


Figure 9: Each diagram shows 10,000 samples of size $N = 7$, with $(\mu, \sigma) = (0, 1)$ and ξ uniform over $[-10, 10]$. To the left are the resulting estimates obtained using Matlab *gevfit*, and to the right using the equal weight combination of elementals.

Statements about the efficacy of ML which are predicated on the parameters lying within some range - such as the regularity condition requiring $\xi < -1/2$ - are somewhat unhelpful, given that the analyst typically does not know the parameters, but has only data. As an example, consider an ordered data set consisting of three points $[-1, x_{\text{mid}}, 1]$ where x_{mid} is some value between -1 and 1. Such data could have originated from a GEV of any ξ , thus questions such as whether $\xi < -1/2$ are not well posed when given only the data. As x_{mid} is varied, the estimates of ξ using the *gevfit* ML algorithm and the (unique) elemental combination are shown in Fig. 10. Again, the irregular graph of the ML estimates is indicative of numerical difficulties, and there were indeed numerous error messages warning of convergence to boundary points or of failure to converge within some pre-designated number of iterations. By contrast, the elemental estimator produces a smooth, well-behaved graph whose computation requires no iteration, being a simple one-step evaluation of the function $\hat{\xi} = a \log \tau - b \log t$ with $a = 1/(3 \log 4/3)$ and $b = 1/(3 \log 3/2)$. Moreover, the resulting graph of the elemental estimate conforms reasonably with intuition, in that the centrally-located $x_{\text{mid}} = 0$ leads to the estimate $\xi = -(\log^2 2)/(3 \log(3/4) \log(2/3)) = -0.23$, which is not very far from the uniform distribution case of $\xi = -1$. (Indeed, there is no reason that the estimate for this uniformly-spaced data case should deliver the exact value $\hat{\xi} = -1$, but some loose agreement is to be expected.) Values of x_{mid} to the right of this correspond to increasingly negative estimates of ξ , which correspond to the data tending to cluster towards the rightmost end point. Similarly, values of x_{mid} further left of centre lead to increasingly positive values of ξ , in correspondence with the tendency of positive ξ distributions to cluster the data to the left end-point. If one assumes that the elemental estimate is in some sense sensible, then it is perhaps surprising to observe that the ML estimate agrees closely with this for large negative values of ξ , the very parameter regime that regularity considerations would have led us to be wary of.

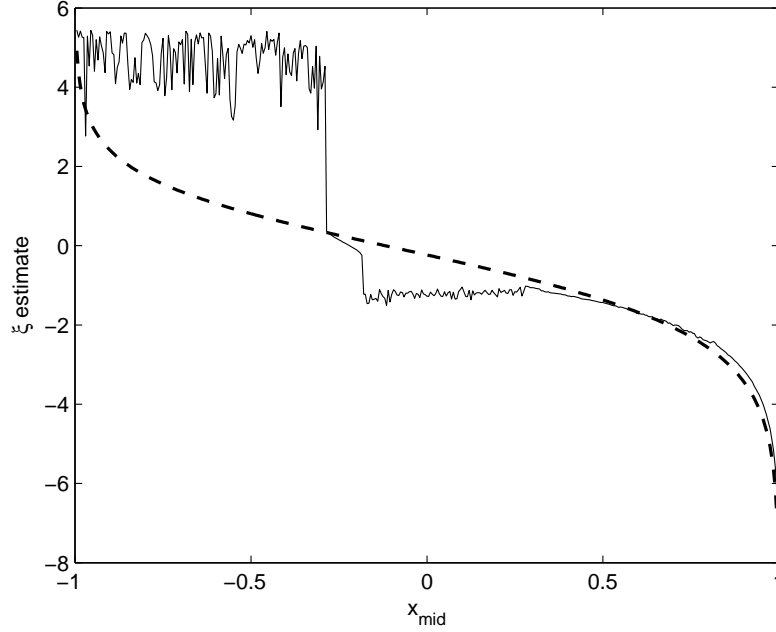


Figure 10: The estimates obtained using Matlab's *gevfit* ML estimator (solid) and the unique elemental (dashed) for the sample $[-1, x_{mid}, 1]$, as x_{mid} is varied.

As further demonstration of the robustness of the elemental approach as compared with the ML approach, the estimator performance is compared over a number of idealised data sets. To create an idealised sample of size N , the unit interval is divided into N equal segments, the midpoints F_i of which are used, via the Probability Integral Transform, to create a sample $\{x_i \mid x_i = F^{-1}(F_i), i = 1, \dots, N\}$ where F is a GEV distribution function with parameters (μ, σ, ξ) with $\mu = 0$, $\sigma = 1$, and ξ varying over the interval $[-10, 10]$. (The ordering of i is such that x_1 is the data maximum). Although each data set employs a nominal value of ξ in its construction, any such data set could have arisen by random sampling from a GEV with any other value of ξ . The value of ξ used in the idealised construction is thus only nominally associated with that data set.

Fig. 11 compares the resulting *gevfit* ML estimates with those obtained from the equal weight elemental combination, for $N = 3, 7, 15$ and 31 . Again, the irregularities of the ML graphs are indicative of numerical difficulties. That these irregularities are present even for $N = 31$ shows that they are not merely a feature of small sample sizes. The elemental estimates however, are well-behaved for all samples, and lie close to the diagonal in all cases. Indeed, given the artificiality of the idealised data construction, although some approximate diagonal correspondence is to be expected, perfect diagonality is not - especially for N small.

In summary, linear combinations of elemental estimators provide computationally efficient and robust estimators for the tail parameter, and - even for extreme parameter values ($|\xi| \approx 10$) and very small sample sizes ($N \approx 3$) - do not suffer the numerical difficulties experienced by the ML estimator.

7 Application to distributions related to Weibull distributions

If the density function of a GEV is flipped right to left, we obtain a three-parameter distribution related to the Weibull distribution. It follows immediately that the shape parameter of a sample drawn from such

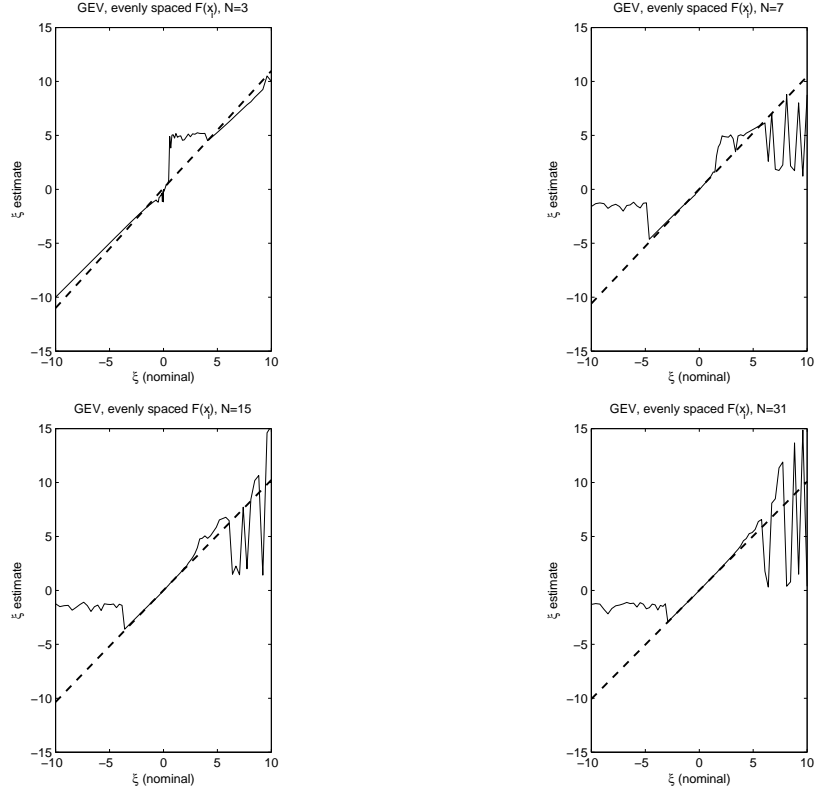


Figure 11: The estimates obtained using the Matlab *gevfit* ML estimator (solid) and the equal weight combination of elementals (dashed) for the idealised samples created by uniformly-spaced F , for $N = 3, 7, 15$ and 31 .

a distribution can be estimated by simply reversing the elemental coefficients. That is, given a sample (ordered such that X_1 is maximum) drawn from the three parameter (μ, σ, ζ) distribution with distribution function

$$F(x) = \begin{cases} 1 - \exp(-(1 - \zeta z)^{-1/\zeta}) & \text{for } \zeta \neq 0 \\ 1 - \exp(-\exp(z)) & \text{for } \zeta = 0 \end{cases} \quad \text{where } z = \frac{x - \mu}{\sigma} \quad (12)$$

an elemental estimator of the parameter ζ can be obtained via

$$\hat{\zeta}_{IJN} = a_N^W(J) \log \tau - b_N^W(I) \log t \quad (13)$$

with τ and t defined as per Eqn. 1 and

$$a_N^W(J) = -b_N(N + 1 - J) \quad \text{and} \quad b_N^W(I) = -a_N(N + 1 - I) \quad (14)$$

For large N , $b_N^W(I)$ converges to $I \log(I/N)$.

The practical significance of all this is that we are often interested in heavy-tailed distributions which head off unbounded to the right, such as the GEV with positive ξ . By flipping the GEV right to left, we obtain another family of heavy-right-tailed distributions (and these are related to the Weibull distribution). These heavy right tails are those that, before flipping, headed off to the left in the GEV for ξ negative. These tails have a different structure to the positive ξ GEV tails, and we now have estimators for the shape parameter of this new class of heavy tails. However the shape parameter ζ of these new tails should not

be confused with their extreme value tail index, the tail parameter ξ of the GEV within whose domain of attraction this Weibull-related distribution lies. That tail index is 0, since the Weibull-related distribution lies within the domain of attraction of the Gumbel distribution, the GEV with $\xi = 0$ (Embrechts et al. (1999)).

A similar reflection could be applied to the GPD estimators of McRobie (2013) but this is of arguably lesser interest, given that all left tails in the GPD are bounded below, and thus when flipped, lead to right tails that are bounded above.

8 Summary

Elemental estimators have been presented for the GEV tail parameter ξ . These have small bias, are computationally simple and robust, are applicable to small data sets and are valid for all parameters ξ , positive, zero and negative. Linear combinations of elementals appear to provide efficient, consistent estimators when applied to data drawn from pure GEVs, and such combinations appear to have advantages over Maximum Likelihood approaches to the GEV. Further, by a simple reflection, the approach has been shown to be applicable to another class of heavy-tailed distributions related to the Weibull distribution.

References

- Castillo, J., Daoudi, J., 2009. Estimation of the generalized Pareto distribution. *Statistics and Probability Letters* 79, 684–688.
- Embrechts, P., Klüppelberg, C., Mikosch, T., 1999. *Modelling Extreme Events for Insurance and Finance*. Springer, Berlin.
- MathWorks, 2012. *Matlab R2012a*.
- McRobie, A., 14 April 2013. Elemental unbiased estimators for the Generalized Pareto tail. arxiv:1304.3918.
- Smith, R. L., 1987. Estimating tails of probability distributions. *Annals of Statistics* 15 (3), 1174–1207.

Appendix 1. Derivation of the elemental coefficients

The Generalised Extreme Value distribution has distribution function

$$F(x) = \begin{cases} \exp \left[-\left(1 + \xi \frac{x-\mu}{\sigma}\right)^{-1/\xi} \right] & \text{for } \xi \neq 0 \\ \exp \left[-\exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right) \right] & \text{for } \xi = 0 \end{cases} \quad (15)$$

The support is $\mu - \sigma/\xi \leq x$ for ξ positive and $x \leq \mu - \sigma/\xi$ for ξ negative. For now, we ignore the $\xi = 0$ case. Inverting the distribution function via the Probability Integral Transform gives

$$x = u(F) = \mu + \frac{\sigma}{\xi} \left[(-\log F)^{-\xi} - 1 \right] \quad (16)$$

Let \mathbf{X} be an ordered sample of size N , ($X_N \leq X_{N-1} \leq \dots \leq X_2 \leq X_1$). The expected value of any function $h(\mathbf{X})$ is then

$$\langle h(\mathbf{X}) \rangle = N! \int_0^1 dF_1 \dots \int_0^{F_{N-1}} dF_N h(u(\mathbf{F})) \quad (17)$$

the integral being over the N -dimensional unit simplex defined by the ordering.

For any function of just two data points X_i and X_j , ($i < j$), all other variables can be integrated out to leave

$$\langle h(X_i, X_j) \rangle = c_{ijN} \int_0^1 dF_i \int_0^{F_i} dF_j f_{ijN}(F_i, F_j) h(u(F_i), u(F_j)) \quad (18)$$

where

$$c_{ijN} = \frac{N!}{(i-1)!(j-i-1)!(N-j)!} \quad \text{and} \quad f_{ijN}(x, y) = (1-x)^{i-1}(x-y)^{j-i-1}y^{N-j} \quad (19)$$

The expected value of any function $h(X_k)$ of a single data point, is likewise, after a further integration, given by

$$\langle h(X_k) \rangle = d_{kN} \int_0^1 dF_k g_{kN}(F_k) h(u(F_k)) \quad (20)$$

where

$$d_{kN} = \frac{N!}{(k-1)!(N-k)!} \quad \text{and} \quad g_{kN}(x) = x^{N-k}(1-x)^{k-1} \quad (21)$$

Expressed in terms of F 's, a log-spacing is given by

$$\log(X_i - X_j) = \begin{cases} \log\left(\frac{\sigma}{\gamma}\right) + \gamma \log(-\log F_j) + \log \left[1 - \left(\frac{-\log F_i}{-\log F_j} \right)^\gamma \right] & \text{for } \xi = -\gamma, \gamma > 0 \\ \log\left(\frac{\sigma}{\gamma}\right) - \gamma \log(-\log F_i) + \log \left[1 - \left(\frac{-\log F_i}{-\log F_j} \right)^\gamma \right] & \text{for } \xi = +\gamma, \gamma > 0 \end{cases} \quad (22)$$

Already, the strategy is becoming evident. The zero sum of log-spacing weights in each elemental will remove the $\log(\sigma/\gamma)$ terms, and expectation of the $\log(-\log F)$ terms will provide some constant multiple of γ that will form the core of the estimator. One hopes that, as in the derivation for the GPD elementals, the remaining complicated terms $\log \left[1 - \left(\frac{-\log F_i}{-\log F_j} \right)^\gamma \right]$ can be combined in such a way as to vanish. Unfortunately this is not the case, although - as will be shown - the residual error is small.

We shall seek an estimator of the form

$$\hat{\xi}_{IJN} = A \log \tau - B \log t \quad (23)$$

with τ and t defined as in Eqn. 1.

For ξ negative ($\xi = -\gamma$, γ positive), we obtain almost immediately

$$\begin{aligned} \langle \log t \rangle_- &= \left\langle \log \left(\frac{X_{I+1} - X_J}{X_I - X_J} \right) \right\rangle \\ &= \left\langle \log \left[1 - \left(\frac{-\log F_{I+1}}{-\log F_J} \right)^\gamma \right] \right\rangle - \left\langle \log \left[1 - \left(\frac{-\log F_I}{-\log F_J} \right)^\gamma \right] \right\rangle \\ &= \int_0^1 dx \int_0^x dy [c_{I+1, JN} f_{I+1, JN}(x, y) - c_{IJN} f_{IJN}(x, y)] \log \left(1 - \left(\frac{-\log x}{-\log y} \right)^\gamma \right) \\ &\equiv I_1 \end{aligned} \quad (24)$$

Essentially there are only the complicated terms, denoted as I_1 .

Similarly, continuing with ξ negative, we obtain

$$\begin{aligned} \langle \log \tau \rangle_- &= \left\langle \log \left(\frac{X_I - X_{J-1}}{X_I - X_J} \right) \right\rangle \\ &= \gamma \langle \log(-\log F_{J-1}) \rangle - \gamma \langle \log(-\log F_J) \rangle \\ &\quad + \left\langle \log \left[1 - \left(\frac{-\log F_I}{-\log F_{J-1}} \right)^\gamma \right] \right\rangle - \left\langle \log \left[1 - \left(\frac{-\log F_I}{-\log F_J} \right)^\gamma \right] \right\rangle \end{aligned}$$

As earlier, the complicated terms can be collected into a single integral, by defining

$$I_2 \equiv \int_0^1 dx \int_0^x dy [c_{I,J-1,N} f_{I,J-1,N}(x,y) - c_{IJN} f_{IJN}(x,y)] \log \left(1 - \left(\frac{-\log x}{-\log y} \right)^\gamma \right) \quad (25)$$

The leading $\log(-\log F)$ terms are each a function of a single data point, and thus via Eqn. 21

$$\langle \log \tau \rangle_- = \gamma \int_0^1 [d_{J-1,N} g_{J-1,N}(x) - d_{JN} g_{JN}(x)] \log(-\log x) dx + I_2 \quad (26)$$

Denoting the term in square brackets as $[DG(x)]$, we have

$$\begin{aligned} [DG] &= \frac{N!}{(J-2)!(N-J+1)!} x^{N-J+1} (1-x)^{J-2} - \frac{N!}{(J-1)!(N-J)!} x^{N-J} (1-x)^{J-1} \\ &= \frac{N!}{(J-1)!(N-J+1)!} [(J-1)x^{N-J+1} (1-x)^{J-2} - (N-J+1)x^{N-J} (1-x)^{J-1}] \end{aligned}$$

Each of the $(1-x)^p$ terms can be expanded using the Binomial Theorem to obtain

$$[DG] = \binom{N}{J-1} \left\{ \sum_{m=0}^{J-2} \frac{(J-1)(J-2)!}{m!(J-2-m)!} (-1)^m x^{N-J+1+m} - \sum_{m=0}^{J-1} \frac{(N-J+1)(J-1)!}{m!(J-1-m)!} (-1)^m x^{N-J+m} \right\} \quad (27)$$

In the first summation, we relabel the dummy summation index m using $k = m + 1$, and then relabel k as m (i.e. m becomes $m - 1$) giving

$$[DG] = \binom{N}{J-1} \left\{ \sum_{m=1}^{J-1} \frac{(J-1)(J-2)!}{(m-1)!(J-1-m)!} (-1)^{m-1} x^{N-J+m} - \sum_{m=0}^{J-1} \frac{(N-J+1)(J-1)!}{m!(J-1-m)!} (-1)^m x^{N-J+m} \right\}$$

Collecting terms for each $m = 1$ to $J - 1$, we obtain

$$[DG] = \binom{N}{J-1} \left\{ -(N-J+1)x^{N-J} - \sum_{m=1}^{J-1} (-1)^m x^{N-J+m} \frac{(J-1)!}{m!(J-1-m)!} (N-J+1+m) \right\} \quad (28)$$

Noting that the leading $-(N-J+1)x^{N-J}$ term is that which would be obtained if the summation were extended down to $m = 0$, we obtain

$$[DG] = - \binom{N}{J-1} \left\{ \sum_{m=0}^{J-1} (-1)^m x^{N-J+m} \binom{J-1}{m} (N-J+1+m) \right\} \quad (29)$$

We now need only multiply by $\log(-\log x)$ and integrate over the unit interval, and use the result that

$$\int_0^1 x^{\alpha-1} \log(-\log x) dx = \frac{-\log \alpha}{\alpha} - \frac{C}{\alpha} \quad (30)$$

where C is Euler's constant.

The terms involving Euler's constant are an alternating sum of binomial coefficients and thus sum to zero:

$$\sum_{m=0}^{J-1} (-1)^m \binom{J-1}{m} = 0 \quad (31)$$

This leaves only the $-(\log \alpha)/\alpha$ terms (where $\alpha = N - J + 1 + m$), which give

$$\int_0^1 [DG(x)] \log(-\log x) dx = \binom{N}{J-1} \left\{ \sum_{m=0}^{J-1} (-1)^m \binom{J-1}{m} \log(N-J+1+m) \right\} \equiv \frac{-1}{A} \quad (32)$$

Collecting all the above, we obtain finally that for ξ negative,

$$\langle \log \tau \rangle_- = \frac{-\gamma}{A} + I_2 = \frac{\xi}{A} + I_2 \quad (33)$$

We now need to repeat the procedure for ξ positive. Writing $\xi = \gamma$, $\gamma > 0$, we readily obtain that

$$\tau = \frac{1 - \left(\frac{-\log F_I}{-\log F_{J-1}} \right)^\gamma}{1 - \left(\frac{-\log F_I}{-\log F_J} \right)^\gamma} \quad (34)$$

whose expected logarithm was defined in Eqn. 25 as I_2 , thus

$$\langle \log \tau \rangle_+ = I_2 \quad (35)$$

Similarly for $\langle \log t \rangle_+$, we obtain an I_1 contribution from the complicated integrals, and the $\log(-\log F)$ terms are

$$\gamma [\langle \log(-\log F_I) \rangle - \langle \log(-\log F_{I+1}) \rangle] \quad (36)$$

These are identical to the terms in the $\langle \log \tau \rangle_-$ derivation, involving integrations over a single variable, except we must make the substitution $J - 1$ goes to I . We thus obtain

$$\langle \log t \rangle_+ = \frac{-\xi}{B} + I_1 \quad (37)$$

with

$$\frac{1}{B} = \int_0^1 [\dots] \log(-\log(x)) dx = - \binom{N}{I} \left\{ \sum_{m=0}^I (-1)^m \binom{I}{m} \log(N - I + m) \right\} \quad (38)$$

In summary,

$$\begin{aligned} \langle \log t \rangle_- &= I_1 \\ \langle \log \tau \rangle_- &= \frac{\xi}{A} + I_2 \\ \langle \log t \rangle_+ &= \frac{-\xi}{B} + I_1 \\ \langle \log \tau \rangle_+ &= I_2 \end{aligned} \quad (39)$$

whence, for positive or negative ξ , we have

$$A \langle \log \tau \rangle - B \langle \log t \rangle = \xi + AI_2 - BI_1 \quad (40)$$

This is an unbiased estimator for ξ provided that the weighted sum of the complicated terms $AI_2 - BI_1 = 0$. Although the equivalent terms in the GPD derivation (McRobie (2013)) cancelled exactly, this is not the case for the GEV here. However the resulting bias is small. It also vanishes as $|\xi|$ becomes large.

Appendix 2. Approximation of the $b_N(I)$ coefficients

$$\text{Writing } \frac{b_N(I)}{N} = \left[N \binom{N}{I} S_I \right]^{-1} \quad \text{with } S_I = - \sum_{m=0}^I \binom{I}{m} (-1)^m \log(N - I + m) \quad (41)$$

$$\text{and } f(x) = -(1-x) \log(1-x) \quad \text{with } x = \frac{I}{N} \quad (42)$$

we show that

$$\frac{b_N(I)}{N} = f(x) \left[1 + \frac{x}{12N} + O\left(\frac{1}{N^3}\right) \right] \quad (43)$$

The demonstration begins by writing the ratio of the approximation $Nf(x)$ to the true value of the coefficient $b_N(I)$ as

$$\frac{f(x)}{b_N(I)/N} = \left[N \binom{N}{I} \right] S_I f(x) \quad (44)$$

and we expand each of the three terms on the right-hand side as a power series in $1/N$, keeping the first three nonzero terms in each expansion.

Term 1:

$$\begin{aligned} N \binom{N}{I} &= \frac{N \cdot N(N-1)(N-2) \dots (N-I+1)}{I!} = \frac{N^{I+1}}{I!} \left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \dots \left(1 - \frac{I-1}{N}\right) \\ &= \frac{N^{I+1}}{I!} \left[1 - \frac{I(I-1)}{2N} + \frac{I(I-1)(I-2)(3I-1)}{24N^2} + O\left(\frac{1}{N^3}\right) \right] \end{aligned} \quad (45)$$

Term 2:

$$\begin{aligned} S_I &= - \sum_{m=0}^I \binom{I}{m} (-1)^m \log(N - I + m) = (-1)^{I+1} \sum_{k=0}^I \binom{I}{k} (-1)^k \log\left(1 - \frac{k}{N}\right) \\ &= (-1)^I \sum_{k=0}^I \binom{I}{k} (-1)^k \sum_{j=1}^{\infty} \frac{1}{j} \left(\frac{k}{N}\right)^j = (-1)^I \sum_{j=1}^{\infty} \frac{1}{jN^j} \sum_{k=0}^I (-1)^k \binom{I}{k} k^j \end{aligned} \quad (46)$$

We now use the relations

$$(-1)^I \sum_{k=0}^I (-1)^k \binom{I}{k} k^j = \begin{cases} 0 & \text{for } j = 1, \dots, I-1 \\ I! & \text{for } j = I \\ [I(I+1)/2] I! & \text{for } j = I+1 \\ [I(I+1)(I+2)(3I+1)/24] I! & \text{for } j = I+2 \end{cases} \quad (47)$$

giving

$$\begin{aligned} S_I &= \frac{I!}{IN^I} + \frac{I(I+1)I!}{2(I+1)N^{I+1}} + \frac{I(I+1)(I+2)(3I+1)I!}{24(I+2)N^{I+2}} + O\left(\frac{1}{N^{I+3}}\right) \\ &= \frac{I!}{IN^I} \left[1 + \frac{I^2}{2N} + \frac{I^2(I+1)(3I+1)}{24N^2} + O\left(\frac{1}{N^3}\right) \right] \end{aligned} \quad (48)$$

This result is worthy of comment. Essentially we have expanded each $\log(1 - k/N)$ as a power series in k/N . When these series are weighted by the binomial coefficients and summed then - rather remarkably - the first $I-1$ terms disappear. For example, if calculating $b_N(I)$ with $I = 100$, the first 99 terms in these series disappear, and the first non-zero terms are those involving $(k/N)^{100}$.

Term 3:

$$\begin{aligned}
f(x) = -(1-x)\log(1-x) &= -(1-x)\left[-x - \frac{x^2}{2} - \frac{x^3}{3} + O(x^4)\right] \\
&= x - \frac{x^2}{2} - \frac{x^3}{6} + O(x^4) = \frac{I}{N}\left[1 - \frac{I}{2N} - \frac{I^2}{6N^2} + O\left(\frac{1}{N^3}\right)\right] \quad (49)
\end{aligned}$$

Combine terms

Each of the three terms is now in the form $c_i(1 + a_i/N + b_i/N^2 + O(1/N^3))$ thus their product can be written

$$\frac{f(x)}{b_N(I)/N} = \left[N \binom{N}{I}\right] S_I f(x) = C \left[1 + \frac{A}{N} + \frac{B}{N^2} + O\left(\frac{1}{N^3}\right)\right] \quad (50)$$

with

$$\begin{aligned}
C &= c_1 c_2 c_3 = \left(\frac{N^{I+1}}{I!}\right) \left(\frac{I!}{IN^I}\right) \left(\frac{I}{N}\right) = 1 \\
A &= a_1 + a_2 + a_3 = \frac{-I(I-1)}{2} + \frac{I^2}{2} + \frac{-I}{2} = 0 \\
\text{and } B &= a_1 a_2 + a_2 a_3 + a_3 a_1 + b_1 + b_2 + b_3 \\
&= \left[\frac{-I(I-1)}{2}\right] \left[\frac{I^2}{2}\right] + \left[\frac{I^2}{2}\right] \left[\frac{-I}{2}\right] + \left[\frac{-I}{2}\right] \left[\frac{-I(I-1)}{2}\right] + \\
&\quad + \frac{I(I-1)(I-2)(3I-1)}{24} + \frac{I^2(I+1)(3I+1)}{24} - \frac{I^2}{6} \\
&= \frac{1}{24} [(-6I^4 + 6I^3) + (-6I^3) + (6I^3 - 6I^2) + (3I^4 - 10I^3 + 9I^2 - 2I) + (3I^4 + 4I^3 + I^2) + (-4I^2)] \\
&= \frac{-I}{12} \quad (51)
\end{aligned}$$

Finally, then

$$\frac{f(x)}{b_N(I)/N} = 1 - \frac{I}{12N^2} + O\left(\frac{1}{N^3}\right) \quad (52)$$

or

$$\frac{b_N(I)}{N} = -(1-x)\log(1-x) \left[1 + \frac{I}{12N^2} + O\left(\frac{1}{N^3}\right)\right] \quad (53)$$

Thus, for I fixed, the coefficient $b_N(I)$ converges to $-N(1-x)\log(1-x)$ as N increases.

The fore-going demonstration would be improved if the convergence were established at fixed x rather than fixed N . However, we note, without further proof, that

$$\frac{b_N(I)}{N} \approx -(1-x)\log(1-x) - \frac{x}{12N}\log(1-x) \quad (54)$$

is a better approximation and this is illustrated in Fig. 4.

The following examples illustrate the somewhat unusual nature of this approximation. Consider the case $N = 20$, $I = 3$.

$$b_{20}(3) = \frac{-3.2.1}{20.19.18} \frac{1}{\log\left(\frac{17.19^3}{18^3.20}\right)} \approx -\left(17 + \frac{3}{12(20)}\right) \log\left(\frac{17}{20}\right) \quad (55)$$

for which a hand-held calculator gives $2.76494701 \approx 2.76494671$. The error is $\approx 3 \times 10^{-7}$, which is substantially better than the $O(1/N^3)$ suggested by the derivation.

Although the approximation was developed for large N , it works surprisingly well even for the smallest possible N (namely $N = 3$), leading to somewhat unusual results such as

$$\begin{aligned} b_3(1) &= \frac{-1}{3 \log(2/3)} \approx -\left(2 + \frac{1}{12(3)}\right) \log(2/3) & \text{or } 0.8221 \approx 0.8222 \\ b_3(2) &= \frac{-1}{3 \log\left(\frac{1(3)}{2^2}\right)} \approx -\left(1 + \frac{2}{12(3)}\right) \log(1/3) & \text{or } 1.1587 \approx 1.1596 \end{aligned}$$

In summary, whilst it is conceptually simple to write down and evaluate the exact expression for $b_N(I)$ in terms of binomially-weighted sums of logarithms, the numerical instabilities are such that the remarkably accurate approximation of Eqn. 54 provides a useful alternative.